

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
21 December 2000 (21.12.2000)

PCT

(10) International Publication Number
WO 00/77202 A1

(51) International Patent Classification⁷: C12N 15/12,
C07K 14/705, C12Q 1/68, G01N 33/50

SmithKline Beecham Pharmaceuticals, New Frontiers
Science Park South, Third Avenue, Harlow, Essex CM19
5AW (GB).

(21) International Application Number: PCT/GB00/02317

(22) International Filing Date: 14 June 2000 (14.06.2000)

(74) Agent: CONNELL, Anthony, Christopher; SmithKline
Beecham, Corporate Intellectual Property, Two New Hori-
zons Court, Brentford, Middlesex TW8 9EP (GB).

(25) Filing Language: English

(81) Designated State (*national*): JP.

(26) Publication Language: English

(84) Designated States (*regional*): European patent (AT, BE,
CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC,
NL, PT, SE).

(30) Priority Data:
9913863.8 15 June 1999 (15.06.1999) GB

(71) Applicant: SMITHKLINE BEECHAM PLC [GB/GB];
New Horizons Court, Brentford, Middlesex TW8 9EP (GB).

Published:
— *With international search report.*
— *Before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments.*

(72) Inventors: KREMER, Andreas; SmithKline Beecham
Pharmaceuticals, New Frontiers Science Park South, Third
Avenue, Harlow, Essex CM19 5AW (GB). JACKSON,
Angela, Rose; SmithKline Beecham Pharmaceuticals,
New Frontiers Science Park South, Third Avenue, Harlow,
Essex CM19 5AW (GB). CAIRNS, William, John;

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*



WO 00/77202 A1

(54) Title: NOVEL COMPOUNDS

(57) Abstract: NOT1a polypeptides and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing NOT1a polypeptides and polynucleotides in therapy, and diagnostic assays for such.

Novel Compounds

Field of the Invention

This invention relates to newly identified polypeptides and polynucleotides encoding such polypeptides, to their use in therapy and in identifying compounds which may be agonists, antagonists and /or inhibitors which are potentially useful in therapy, and to production of such polypeptides and polynucleotides.

Background of the Invention

The drug discovery process is currently undergoing a fundamental revolution as it embraces 'functional genomics', that is, high throughput genome- or gene-based biology. This approach as a means to identify genes and gene products as therapeutic targets is rapidly superceding earlier approaches based on 'positional cloning'. A phenotype, that is a biological function or genetic disease, would be identified and this would then be tracked back to the responsible gene, based on its genetic map position.

Functional genomics relies heavily on high-throughput DNA sequencing technologies and the various tools of bioinformatics to identify gene sequences of potential interest from the many molecular biology databases now available. There is a continuing need to identify and characterise further genes and their related polypeptides/proteins, as targets for drug discovery.

Summary of the Invention

The present invention relates to the nuclear receptor NOT1 splice variant, NOT1a, in particular NOT1a polypeptides and NOT1a polynucleotides, recombinant materials and methods for their production. In another aspect, the invention relates to methods for using such polypeptides and polynucleotides, including the treatment of disease states that involve T-cell function, metabolic disorders, cancer and neurological diseases. In a further aspect, the invention relates to methods for identifying agonists and antagonists/inhibitors using the materials provided by the invention, and treating conditions associated with NOT1a imbalance with the identified compounds. In a still further aspect, the invention relates to diagnostic assays for detecting diseases associated with inappropriate NOT1a activity or levels.

30

Description of the Invention

In a first aspect, the present invention relates to NOT1a polypeptides. Such peptides include isolated polypeptides comprising an amino acid sequence which has at least 95% identity, preferably at least 97-99% identity, to that of SEQ ID NO:2 over the entire length of SEQ ID NO:2. Such polypeptides include those comprising the amino acid of SEQ ID NO:2.

Further peptides of the present invention include isolated polypeptides in which the amino acid sequence has at least 95% identity, preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:2 over the entire length of SEQ ID NO:2. Such polypeptides include the polypeptide of SEQ ID NO:2.

5 Further peptides of the present invention include isolated polypeptides encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:1.

Polypeptides of the present invention are members of the nuclear receptor family of polypeptides. They are therefore of interest because members of this family have been shown to directly regulate the expression of genes at a transcriptional level, and a large number of drugs on
10 the market today target certain nuclear receptors for therapeutic benefit. Multiple human diseases including neurological and immune disorders, cancer and cardiovascular disease result from imbalances in the control of gene expression. Therefore, with nuclear receptors being both transcription factors and ligand regulated, they represent an ideal class of drug target for correcting the aberrant control of gene expression. The NOT1 subfamily of nuclear receptors has
15 been demonstrated to be involved in T-cell function, hypothalamus-pituitary-axis control and in neuronal cell survival. Hence, the ability to control NOT1a activity, for example, by developing agonists and antagonists of its function, would be of benefit in various disease states that involve the above mentioned systems. These properties are hereinafter referred to as "NOT1a activity" or "NOT1a polypeptide activity" or "biological activity of human NOT1a". Also included amongst
20 these activities are antigenic and immunogenic activities of said NOT1a polypeptides, in particular the antigenic and immunogenic activities of the polypeptide of SEQ ID NO:2. Preferably, a polypeptide of the present invention exhibits at least one biological activity of human NOT1a.

The polypeptides of the present invention may be in the form of the "mature" protein or may be a part of a larger protein such as a precursor or a fusion protein. It is often advantageous
25 to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The present invention also includes variants of the aforementioned polypeptides, that is polypeptides that vary from the referents by conservative amino acid substitutions, whereby a residue
30 is substituted by another with like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5-10, 1-5, 1-3, 1-2 or 1 amino acids are substituted, deleted, or added in any combination.

35 Polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced

polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

In a further aspect, the present invention relates to NOT1 polynucleotides. Such polynucleotides include isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide which has at least 95% identity to the amino acid sequence of SEQ ID NO:2, over the entire length of SEQ ID NO:2. In this regard, polypeptides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising the nucleotide sequence contained in SEQ ID NO:1 encoding the polypeptide of SEQ ID NO:2.

Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence that has at least 95% identity to a nucleotide sequence encoding a polypeptide of SEQ ID NO:2, over the entire coding region. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred.

Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence which has at least 95% identity to SEQ ID NO:1 over the entire length of SEQ ID NO:1. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising the polynucleotide of SEQ ID NO:1 as well as the polynucleotide of SEQ ID NO:1.

The invention also provides polynucleotides which are complementary to all the above described polynucleotides.

The nucleotide sequence of SEQ ID NO:1 is a splice variant of human NOT1 which lacks part of exon 7 (missing nucleotides 1362-1482). The full-length nucleotide sequence of human NOT1 is given in SEQ ID NO:3. SEQ ID NO:3 is a cDNA sequence and comprises a polypeptide encoding sequence (nucleotide 1-1797 including stop codon), coding exons, exon 3(coding nucleotides 1-864), exon 4(865-994), Exon 5(995-1158), Exon 6(1159-1361), Exon 7(1362-1540), Exon 8(1541-1797), encoding a polypeptide of 598 amino acids, the polypeptide of SEQ ID NO:4.

SEQ ID NO:1 is a cDNA sequence and comprises a polypeptide encoding sequence (nucleotide 1-1676) encoding a polypeptide of 455 amino acids, the polypeptide of SEQ ID NO:2. The nucleotide sequence encoding the polypeptide of SEQ ID NO:2 may be identical to the polypeptide encoding sequence contained in SEQ ID NO:1 or it may be a sequence other than the one contained in SEQ ID NO:1, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:2.

Preferred polypeptides and polynucleotides of the present invention are expected to have, *inter alia*, similar biological functions/properties to their homologous polypeptides and polynucleotides. Furthermore, preferred polypeptides and polynucleotides of the present invention have at least one NOT1a activity.

5 Polynucleotides of the present invention may be obtained, using standard cloning and screening techniques, from a cDNA library derived from mRNA in cells of human fetal brain, (Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.(1989)). Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well
10 known and commercially available techniques.

When polynucleotides of the present invention are used for the recombinant production of polypeptides of the present invention, the polynucleotide may include the coding sequence for the mature polypeptide, by itself, or the coding sequence for the mature polypeptide in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or
15 prepro- protein sequence, or other fusion peptide portions. For example, a marker sequence which facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, Proc Natl Acad Sci USA (1989) 86:821-824, or is an HA tag. The polynucleotide may also contain non-coding 5' and 3' sequences, such as
20 transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Further embodiments of the present invention include polynucleotides encoding polypeptide variants which comprise the amino acid sequence of SEQ ID NO:2 and in which several, for instance from 5 to 10, 1 to 5, 1 to 3, 1 to 2 or 1, amino acid residues are substituted, deleted or added, in any
25 combination.

Polynucleotides which are identical or sufficiently identical to a nucleotide sequence contained in SEQ ID NO:1, may be used as hybridization probes for cDNA and genomic DNA or as primers for a nucleic acid amplification (PCR) reaction, to isolate full-length cDNAs and genomic clones encoding polypeptides of the present invention and to isolate cDNA and genomic clones of
30 other genes (including genes encoding paralogs from human sources and orthologs and paralogs from species other than human) that have a high sequence similarity to SEQ ID NO:1. Typically these nucleotide sequences are 70% identical, preferably 80% identical, more preferably 90% identical, most preferably 95% identical to that of the referent. The probes or primers will generally comprise at least 15 nucleotides, preferably, at least 30 nucleotides and may have at least 50
35 nucleotides. Particularly preferred probes will have between 30 and 50 nucleotides. Particularly preferred primers will have between 20 and 25 nucleotides.

A polynucleotide encoding a polypeptide of the present invention, including homologs from species other than human, may be obtained by a process which comprises the steps of screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof; and isolating full-length cDNA and genomic clones
5 containing said polynucleotide sequence. Such hybridization techniques are well known to the skilled artisan. Preferred stringent hybridization conditions include overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA; followed by washing the filters in 0.1x SSC at about 65°C.
10 Thus the present invention also includes polynucleotides obtainable by screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 or a fragment thereof.

The skilled artisan will appreciate that, in many cases, an isolated cDNA sequence will be incomplete, in that the region coding for the polypeptide is short at the 5' end of the cDNA. This
15 is a consequence of reverse transcriptase, an enzyme with inherently low 'processivity' (a measure of the ability of the enzyme to remain attached to the template during the polymerisation reaction), failing to complete a DNA copy of the mRNA template during 1st strand cDNA synthesis.

There are several methods available and well known to those skilled in the art to obtain
20 full-length cDNAs, or extend short cDNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, for example, Frohman et al., PNAS USA 85, 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon™ technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon™ technology, cDNAs have been prepared from mRNA extracted from
25 a chosen tissue and an 'adaptor' sequence ligated onto each end. Nucleic acid amplification (PCR) is then carried out to amplify the 'missing' 5' end of the cDNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using 'nested' primers, that is, primers designed to anneal within the amplified product (typically an adaptor specific primer that anneals further 3' in the adaptor sequence and a gene specific primer
30 that anneals further 5' in the known gene sequence). The products of this reaction can then be analysed by DNA sequencing and a full-length cDNA constructed either by joining the product directly to the existing cDNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the 5' primer.

Recombinant polypeptides of the present invention may be prepared by processes well
35 known in the art from genetically engineered host cells comprising expression systems. Accordingly,

in a further aspect, the present invention relates to expression systems which comprise a polynucleotide or polynucleotides of the present invention, to host cells which are genetically engineered with such expression systems and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Introduction of polynucleotides into host cells can be effected by methods described in many standard laboratory manuals, such as Davis et al., *Basic Methods in Molecular Biology* (1986) and Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). Preferred such methods include, for instance, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

Representative examples of appropriate hosts include bacterial cells, such as *Streptococci*, *Staphylococci*, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used, for instance, chromosomal, episomal and virus-derived systems, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector which is able to maintain, propagate or express a polynucleotide to produce a polypeptide in a host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (supra). Appropriate secretion signals may be incorporated into the desired polypeptide to allow secretion of the translated protein into the lumen of the endoplasmic reticulum, the periplasmic space or the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals.

If a polypeptide of the present invention is to be expressed for use in screening assays, it is generally preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If the polypeptide is secreted into the

medium, the medium can be recovered in order to recover and purify the polypeptide. If produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

Polypeptides of the present invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid
5 extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active
10 conformation when the polypeptide is denatured during intracellular synthesis, isolation and or purification.

This invention also relates to the use of polynucleotides of the present invention as diagnostic reagents. Detection of a mutated form of the gene characterised by the polynucleotide of SEQ ID NO:1 which is associated with a dysfunction will provide a diagnostic tool that can add to,
15 or define, a diagnosis of a disease, or susceptibility to a disease, which results from under-expression, over-expression or altered spatial or temporal expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification techniques prior to analysis.
20 RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled NOT1a nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations
25 in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (see, e.g., Myers *et al.*, Science (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton *et al.*, Proc Natl Acad Sci USA (1985) 85: 4397-4401). In another embodiment, an array of oligonucleotides probes comprising NOT1a nucleotide sequence or
30 fragments thereof can be constructed to conduct efficient screening of e.g., genetic mutations. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M.Chee *et al.*, Science, Vol 274, pp 610-613 (1996)).

The diagnostic assays offer a process for diagnosing or determining a susceptibility to the
35 Diseases through detection of mutation in the NOT1a gene by the methods described. In addition, such diseases may be diagnosed by methods comprising determining from a sample derived from

a subject an abnormally decreased or increased level of polypeptide or mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

- 5 Assay techniques that can be used to determine levels of a protein, such as a polypeptide of the present invention, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Thus in another aspect, the present invention relates to a diagnostic kit which comprises:

- 10 (a) a polynucleotide of the present invention, preferably the nucleotide sequence of SEQ ID NO: 1, or a fragment thereof ;
(b) a nucleotide sequence complementary to that of (a);
(c) a polypeptide of the present invention, preferably the polypeptide of SEQ ID NO:2 or a fragment thereof; or
15 (d) an antibody to a polypeptide of the present invention, preferably to the polypeptide of SEQ ID NO:2.

- It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component. Such a kit will be of use in diagnosing a disease or susceptibility to a disease, particularly disease states that involve T-cell function, metabolic disorders, cancer and neurological
20 diseases.

- The nucleotide sequences of the present invention are also valuable for chromosomal localisation. The sequence is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important first step in correlating those sequences with gene associated
25 disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).
30

- The differences in the cDNA or genomic sequence between affected and unaffected individuals can also be determined. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

- 35 The gene of the present invention maps to human chromosome 2q22-23.

The nucleotide sequences of the present invention are also valuable for tissue localisation. Such techniques allow the determination of expression patterns of the human NOT1a polypeptides in tissues by detection of the mRNAs that encode them. These techniques include in situ hybridization techniques and nucleotide amplification techniques, for example PCR. Such techniques are well known in the art. Results from these studies provide an indication of the normal functions of the polypeptides in the organism. In addition, comparative studies of the normal expression pattern of human NOT1a mRNAs with that of mRNAs encoded by a human NOT1a gene provide valuable insights into the role of mutant human NOT1a polypeptides, or that of inappropriate expression of normal human NOT1a polypeptides, in disease. Such inappropriate expression may be of a temporal, spatial or simply quantitative nature.

The polypeptides of the invention or their fragments or analogs thereof, or cells expressing them, can also be used as immunogens to produce antibodies immunospecific for polypeptides of the present invention. The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art.

Antibodies generated against polypeptides of the present invention may be obtained by administering the polypeptides or epitope-bearing fragments, analogs or cells to an animal, preferably a non-human animal, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., *Nature* (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today* (1983) 4:72) and the EBV-hybridoma technique (Cole *et al.*, *Monoclonal Antibodies and Cancer Therapy*, 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies, such as those described in U.S. Patent No. 4,946,778, can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms, including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography.

Antibodies against polypeptides of the present invention may also be employed to treat the Diseases, amongst others.

In a further aspect, the present invention relates to genetically engineered soluble fusion proteins comprising a polypeptide of the present invention, or a fragment thereof, and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclasses (IgG, IgM, IgA, IgE). Preferred as an immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. In a

particular embodiment, the Fc part can be removed simply by incorporation of a cleavage sequence which can be cleaved with blood clotting factor Xa. Furthermore, this invention relates to processes for the preparation of these fusion proteins by genetic engineering, and to the use thereof for drug screening, diagnosis and therapy. A further aspect of the invention also relates to
5 polynucleotides encoding such fusion proteins. Examples of fusion protein technology can be found in International Patent Application Nos. WO94/29458 and WO94/22914.

Another aspect of the invention relates to a method for inducing an immunological response in a mammal which comprises inoculating the mammal with a polypeptide of the present invention, adequate to produce antibody and/or T cell immune response to protect said animal
10 from the Diseases hereinbefore mentioned, amongst others. Yet another aspect of the invention relates to a method of inducing immunological response in a mammal which comprises, delivering a polypeptide of the present invention *via* a vector directing expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases.

15 A further aspect of the invention relates to an immunological/vaccine formulation (composition) which, when introduced into a mammalian host, induces an immunological response in that mammal to a polypeptide of the present invention wherein the composition comprises a polypeptide or polynucleotide of the present invention. The vaccine formulation may further comprise a suitable carrier. Since a polypeptide may be broken down in the stomach, it is
20 preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The
25 formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of
30 the vaccine and can be readily determined by routine experimentation.

Polypeptides of the present invention are responsible for one or more biological functions, including one or more disease states, in particular the Diseases hereinbefore mentioned. It is therefore desirous to devise screening methods to identify compounds which stimulate or which inhibit the function of the polypeptide. Accordingly, in a further aspect, the present invention
35 provides for a method of screening compounds to identify those which stimulate or which inhibit the function of the polypeptide. In general, agonists or antagonists may be employed for therapeutic and

prophylactic purposes for such Diseases as hereinbefore mentioned. Compounds may be identified from a variety of sources, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. Such agonists, antagonists or inhibitors so-identified may be natural or modified substrates, ligands, receptors, enzymes, etc., as the case may be, of the polypeptide; or may be structural or functional mimetics thereof (see Coligan *et al.*, Current Protocols in Immunology 1(2):Chapter 5 (1991)).

The screening method may simply measure the binding of a candidate compound to the polypeptide, or to cells or membranes bearing the polypeptide, or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound. Alternatively, the screening method may involve competition with a labeled competitor. Further, these screening methods may test whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells bearing the polypeptide. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed. Constitutively active polypeptides may be employed in screening methods for inverse agonists or inhibitors, in the absence of an agonist or inhibitor, by testing whether the candidate compound results in inhibition of activation of the polypeptide. Further, the screening methods may simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide of the present invention, to form a mixture, measuring NOT1a activity in the mixture, and comparing the NOT1a activity of the mixture to a standard. Fusion proteins, such as those made from Fc portion and NOT1a polypeptide, as hereinbefore described, can also be used for high-throughput screening assays to identify antagonists for the polypeptide of the present invention (see D. Bennett *et al.*, J Mol Recognition, 8:52-58 (1995); and K. Johanson *et al.*, J Biol Chem, 270(16):9459-9471 (1995)).

The polynucleotides, polypeptides and antibodies to the polypeptide of the present invention may also be used to configure screening methods for detecting the effect of added compounds on the production of mRNA and polypeptide in cells. For example, an ELISA assay may be constructed for measuring secreted or cell associated levels of polypeptide using monoclonal and polyclonal antibodies by standard methods known in the art. This can be used to discover agents which may inhibit or enhance the production of polypeptide (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues.

The polypeptide may be used to identify membrane bound or soluble receptors, if any, through standard receptor binding techniques known in the art. These include, but are not limited to, ligand binding and crosslinking assays in which the polypeptide is labeled with a radioactive isotope (for instance, ^{125}I), chemically modified (for instance, biotinylated), or fused to a peptide sequence suitable for detection or purification, and incubated with a source of the putative

receptor (cells, cell membranes, cell supernatants, tissue extracts, bodily fluids). Other methods include biophysical techniques such as surface plasmon resonance and spectroscopy. These screening methods may also be used to identify agonists and antagonists of the polypeptide which compete with the binding of the polypeptide to its receptors, if any. Standard methods for

5 conducting such assays are well understood in the art.

Examples of potential polypeptide antagonists include antibodies or, in some cases, oligonucleotides or proteins which are closely related to the ligands, substrates, receptors, enzymes, etc., as the case may be, of the polypeptide, e.g., a fragment of the ligands, substrates, receptors, enzymes, etc.; or small molecules which bind to the polypeptide of the present invention but do not

10 elicit a response, so that the activity of the polypeptide is prevented.

Thus, in another aspect, the present invention relates to a screening kit for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, etc. for polypeptides of the present invention; or compounds which decrease or enhance the production of such polypeptides, which comprises:

- 15 (a) a polypeptide of the present invention;
(b) a recombinant cell expressing a polypeptide of the present invention;
(c) a cell membrane expressing a polypeptide of the present invention; or
(d) antibody to a polypeptide of the present invention;
which polypeptide is preferably that of SEQ ID NO:2.

20 It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component.

It will be readily appreciated by the skilled artisan that a polypeptide of the present invention may also be used in a method for the structure-based design of an agonist, antagonist or inhibitor of the polypeptide, by:

- 25 (a) determining in the first instance the three-dimensional structure of the polypeptide;
(b) deducing the three-dimensional structure for the likely reactive or binding site(s) of an agonist, antagonist or inhibitor;
(c) synthesizing candidate compounds that are predicted to bind to or react with the deduced binding or reactive site; and
30 (d) testing whether the candidate compounds are indeed agonists, antagonists or inhibitors.

It will be further appreciated that this will normally be an iterative process.

In a further aspect, the present invention provides methods of treating abnormal conditions such as, for instance, disease states that involve T-cell function, metabolic disorders, cancer and neurological diseases, related to either an excess of, or an under-expression of, NOT1a polypeptide

35 activity.

If the activity of the polypeptide is in excess, several approaches are available. One approach comprises administering to a subject in need thereof an inhibitor compound (antagonist) as hereinabove described, optionally in combination with a pharmaceutically acceptable carrier, in an amount effective to inhibit the function of the polypeptide, such as, for example, by blocking the binding of ligands, substrates, receptors, enzymes, etc., or by inhibiting a second signal, and thereby alleviating the abnormal condition. In another approach, soluble forms of the polypeptides still capable of binding the ligand, substrate, enzymes, receptors, etc. in competition with endogenous polypeptide may be administered. Typical examples of such competitors include fragments of the NOT1a polypeptide.

In still another approach, expression of the gene encoding endogenous NOT1a polypeptide can be inhibited using expression blocking techniques. Known such techniques involve the use of antisense sequences, either internally generated or externally administered (see, for example, O'Connor, J Neurochem (1991) 56:560 in Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Alternatively, oligonucleotides which form triple helices ("triplexes") with the gene can be supplied (see, for example, Lee *et al.*, Nucleic Acids Res (1979) 6:3073; Cooney *et al.*, Science (1988) 241:456; Dervan *et al.*, Science (1991) 251:1360). These oligomers can be administered *per se* or the relevant oligomers can be expressed *in vivo*. Synthetic antisense or triplex oligonucleotides may comprise modified bases or modified backbones. Examples of the latter include methylphosphonate, phosphorothioate or peptide nucleic acid backbones. Such backbones are incorporated in the antisense or triplex oligonucleotide in order to provide protection from degradation by nucleases and are well known in the art. Antisense and triplex molecules synthesised with these or other modified backbones also form part of the present invention.

In addition, expression of the human NOT1a polypeptide may be prevented by using ribozymes specific to the human NOT1a mRNA sequence. Ribozymes are catalytically active RNAs that can be natural or synthetic (see for example Usman, N, et al., Curr. Opin. Struct. Biol (1996) 6(4), 527-33.) Synthetic ribozymes can be designed to specifically cleave human NOT1a mRNAs at selected positions thereby preventing translation of the human NOT1a mRNAs into functional polypeptide. Ribozymes may be synthesised with a natural ribose phosphate backbone and natural bases, as normally found in RNA molecules. Alternatively the ribozymes may be synthesised with non-natural backbones to provide protection from ribonuclease degradation, for example, 2'-O-methyl RNA, and may contain modified bases.

For treating abnormal conditions related to an under-expression of NOT1a and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound which activates a polypeptide of the present invention, i.e., an agonist as described above, in combination with a pharmaceutically acceptable

carrier, to thereby alleviate the abnormal condition. Alternatively, gene therapy may be employed to effect the endogenous production of NOT1a by the relevant cells in the subject. For example, a polynucleotide of the invention may be engineered for expression in a replication defective retroviral vector, as discussed above. The retroviral expression construct may then be isolated and introduced
5 into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells *in vivo* and expression of the polypeptide *in vivo*. For an overview of gene therapy, see Chapter 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (and
10 references cited therein) in Human Molecular Genetics, T Strachan and A P Read, BIOS Scientific Publishers Ltd (1996). Another approach is to administer a therapeutic amount of a polypeptide of the present invention in combination with a suitable pharmaceutical carrier.

In a further aspect, the present invention provides for pharmaceutical compositions comprising a therapeutically effective amount of a polypeptide, such as the soluble form of a
15 polypeptide of the present invention, agonist/antagonist peptide or small molecule compound, in combination with a pharmaceutically acceptable carrier or excipient. Such carriers include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.
20 Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The composition will be adapted to the route of administration, for instance by a systemic or an oral route. Preferred forms of systemic administration include injection, typically by intravenous injection. Other injection routes, such as subcutaneous, intramuscular, or intraperitoneal, can be
25 used. Alternative means for systemic administration include transmucosal and transdermal administration using penetrants such as bile salts or fusidic acids or other detergents. In addition, if a polypeptide or other compounds of the present invention can be formulated in an enteric or an encapsulated formulation, oral administration may also be possible. Administration of these compounds may also be topical and/or localized, in the form of salves, pastes, gels, and the like.

30 The dosage range required depends on the choice of peptide or other compounds of the present invention, the route of administration, the nature of the formulation, the nature of the subject's condition, and the judgment of the attending practitioner. Suitable dosages, however, are in the range of 0.1-100 µg/kg of subject. Wide variations in the needed dosage, however, are to be expected in view of the variety of compounds available and the differing efficiencies of various
35 routes of administration. For example, oral administration would be expected to require higher

dosages than administration by intravenous injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, as is well understood in the art.

Polypeptides used in treatment can also be generated endogenously in the subject, in treatment modalities often referred to as "gene therapy" as described above. Thus, for example, cells from a subject may be engineered with a polynucleotide, such as a DNA or RNA, to encode a polypeptide *ex vivo*, and for example, by the use of a retroviral plasmid vector. The cells are then introduced into the subject.

Polynucleotide and polypeptide sequences form a valuable information resource with which to identify further sequences of similar homology. This is most easily facilitated by storing the sequence in a computer readable medium and then using the stored data to search a sequence database using well known searching tools, such as those in GCG and Lasergene software packages. Accordingly, in a further aspect, the present invention provides for a computer readable medium having stored thereon a polynucleotide comprising the sequence of SEQ ID NO:1 and/or a polypeptide sequence encoded thereby.

The following definitions are provided to facilitate understanding of certain terms used frequently hereinbefore.

"Antibodies" as used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the products of an Fab or other immunoglobulin expression library.

"Isolated" means altered "by the hand of man" from the natural state. If an "isolated" composition or substance occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term "polynucleotide" also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons.

"Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications may be made to DNA and RNA; thus, "polynucleotide" embraces

chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells.

"Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

5 "Polypeptide" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either
10 by natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications may occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may
15 be present to the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation,
20 biotinylation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation,
25 hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination (see, for instance, *Proteins - Structure and Molecular Properties*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993; Wold, F., *Post-translational Protein Modifications: Perspectives and Prospects*, pgs. 1-12 in *Post-translational Covalent Modification of Proteins*, B. C. Johnson, Ed.,
30 Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", *Meth Enzymol* (1990) 182:626-646 and Rattan *et al.*, "Protein Synthesis: Post-translational Modifications and Aging", *Ann NY Acad Sci* (1992) 663:48-62).

 "Variant" refers to a polynucleotide or polypeptide that differs from a reference
35 polynucleotide or polypeptide, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in

the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., *SIAM J. Applied Math.*, 48: 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S.F. et al., *J. Molec. Biol.* 215: 403-410 (1990)). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., *J. Mol. Biol.* 215: 403-410 (1990)). The well known Smith Waterman algorithm may also be used to determine identity.

Preferred parameters for polypeptide sequence comparison include the following:

- 1) Algorithm: Needleman and Wunsch, *J. Mol Biol.* 48: 443-453 (1970)

Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, Proc. Natl. Acad. Sci. USA. 89:10915-10919 (1992)

Gap Penalty: 12

Gap Length Penalty: 4

- 5 A program useful with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for peptide comparisons (along with no penalty for end gaps).

Preferred parameters for polynucleotide comparison include the following:

1) Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)

- 10 Comparison matrix: matches = +10, mismatch = 0

Gap Penalty: 50

Gap Length Penalty: 3

Available as: The "gap" program from Genetics Computer Group, Madison WI. These are the default parameters for nucleic acid comparisons.

- 15 By way of example, a polynucleotide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:1, that is be 100% identical, or it may include up to a certain integer number of nucleotide alterations as compared to the reference sequence. Such alterations are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. The number of nucleotide alterations is determined by multiplying the total number of nucleotides in SEQ ID NO:1 by the numerical percent of the respective percent identity (divided by 100) and subtracting that product from said total number of nucleotides in SEQ ID NO:1, or:
- 20
- 25

$$n_n \leq x_n - (x_n \cdot y),$$

wherein n_n is the number of nucleotide alterations, x_n is the total number of nucleotides in SEQ ID NO:1, and y is, for instance, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%, etc., and wherein any non-integer product of x_n and y is rounded down to the nearest integer prior to subtracting it from x_n . Alterations of a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 may create nonsense, missense or frameshift mutations in this coding sequence and thereby alter the polypeptide encoded by the polynucleotide following such alterations.

30

- 35 Similarly, a polypeptide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:2, that is be 100% identical, or it may include up to a certain

integer number of amino acid alterations as compared to the reference sequence such that the % identity is less than 100%. Such alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of amino acid alterations for a given % identity is determined by multiplying the total number of amino acids in SEQ ID NO:2 by the numerical percent of the respective percent identity (divided by 100) and then subtracting that product from said total number of amino acids in SEQ ID NO:2, or:

$$n_a \leq x_a - (x_a \cdot y),$$

wherein n_a is the number of amino acid alterations, x_a is the total number of amino acids in SEQ ID NO:2, and y is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., and wherein any non-integer product of x_a and y is rounded down to the nearest integer prior to subtracting it from x_a .

"Homolog" is a generic term used in the art to indicate a polynucleotide or polypeptide sequence possessing a high degree of sequence relatedness to a subject sequence. Such relatedness may be quantified by determining the degree of identity and/or similarity between the sequences being compared as hereinbefore described. Falling within this generic term are the terms "ortholog", meaning a polynucleotide or polypeptide that is the functional equivalent of a polynucleotide or polypeptide in another species, and "paralog" meaning a functionally similar sequence when considered within the same species.

"Fusion protein" refers to a protein encoded by two, often unrelated, fused genes or fragments thereof. In one example, EP-A-0 464 discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, employing an immunoglobulin Fc region as a part of a fusion protein is advantageous for use in therapy and diagnosis resulting in, for example, improved pharmacokinetic properties [see, e.g., EP-A 0232 262]. On the other hand, for some uses it would be desirable to be able to delete the Fc part after the fusion protein has been expressed, detected and purified.

Examples**Example 1: Comparison of NOT1 and NOT1a expression in human tissues**

Generation of samples for TaqMan mRNA analysis: Human tissue or RNA was purchased (Biochain, San Leandro, CA; Invitrogen, Leek, The Netherlands; Clontech, Palo Alto, CA) or donated (Netherlands Brain Bank, Amsterdam, the Netherlands) and poly A+ RNA was prepared by the PolyAtract method according to manufacturers instructions (Promega, USA). The poly A+ RNA samples from 20 body tissues and 19 brain-regions from 4 individuals per tissue (two males/two females) were quantitated using OD260nm measurement or the RiboGreen fluorescent method (Molecular Probes, Oregon, USA) and 1 µg of each RNA was reverse transcribed using random nonomers and Superscript II reverse transcriptase according to manufacturers instructions (Life Technologies). The cDNA prepared was diluted to produce up to 1,000 replicate 96-well plates using Biomek robotics (Beckman Coulter, High Wycombe, UK), so that each of the wells contained the cDNA produced from 1 ng RNA for the appropriate tissue. The 96-well plates were stored at -80°C prior to use.

TaqMan PCR: This was performed following the procedure published by Sarau H.M. et al, (1999) Molecular Pharmacology 56, 657-663). TaqMan quantitative PCR was conducted to measure either NOT-1 or NOT1a using replicate 96-well plates. A 20 µl volume of a PCR master mix (containing 2.5 µl TaqMan buffer, 6 µl 25 mM MgCl₂, 0.5 µl of 10 mM dATP, 0.5 µl of 20 mM dUTP, 0.5 µl of 10 mM dCTP, 0.5 µl of 10 mM dGTP, 0.25 µl Uracil-N-glycosylase, 1 µl of 10 uM forward primer, 1 µl of 10 uM reverse primer, 0.5 µl 5 uM TaqMan probe, 0.125 µl TaqGold [PE Biosystems], 6.625 µl water) was added to each well using Biomek robotics (Beckman Coulter, High Wycombe, UK), and the plate capped using optical caps (PE Biosystems). The PCR reaction was carried out on an ABI7700 Sequence Detector (PE Biosystems) using the PCR parameters: 50°C for 2 minutes, 95°C for 10 minutes and 45 cycles of 94°C for 15 seconds, 60°C for 1 minute, and the level of mRNA-derived cDNA in each sample was calculated from the TaqMan signal using plasmid/genomic DNA calibration standards included in each run. The level of genomic DNA contaminating the original RNA samples was shown to be negligible (<10 copies genomic DNA/ng RNA) by TaqMan measurement of genomic sequence for ten genes in replicate samples taken through the reverse transcription procedure described with the omission of reverse transcriptase. Gene-specific reagents for NOT-1:

NOT-1 forward primer: 5'-CCGCCAGCAATAATTGACAA-3';

NOT-1 reverse primer: 5'-TTCCATTATCATTCCAGTTCCTTTG-3';

NOT-1 TaqMan probe: 5'-CACTTTACCTTTCTAAGACCTCCTCCCAAGCA-3';

NOT1a forward primer: 5'-TCCTTCGATTAGCATACAGAATA-3';

NOT-1a reverse primer: 5'-CCCGTGTCTCTCTGTGACCAT-3';

NOT-1a TaqMan probe: 5'-TCTGCCTTCTCCTGCATTGCTGCC-3'.

5 Results

Analysis of expression of both NOT1 and NOT1a mRNA by TaqMan showed a relatively widespread distribution. Although the expression levels of NOT1a were generally lower than for full length NOT1, their patterns of distribution were similar.

	NOT1	NOT1a
10 Brain	++	+
Pituitary	+++	++
Heart	+	+/-
Lung	+	+/-
15 Liver	+	+
Foetal Liver	+	+/-
Kidney	+	+
Skeletal Muscle	++	++
Stomach	+	+
20 Intestine	+	+
Spleen	+	+
Lymphocytes	++	++
Macrophage	-	-
Adipose	+++++	+++++
25 Pancreas	+	+
Prostate	++	+
Placenta	+	+
Cartilage	+	+
Bone	-	-
30 Bone Marrow	+/-	+/-

Example 2: Comparison of NOT1 and NOT1a transcriptional activity.

35 Vector constructs: *NOT1 reporter*- Two copies of the NOT1 DNA response element:

5'-TCGAGATCGCAGGTCACCGAAAGGTCATGAATC-3' were subcloned upstream of the thymidine kinase promoter driving luciferase expression.

NOT1 and NOT1a expression constructs- NOT1 and NOT1a cDNAs were subcloned into the expression vector pcDNA3.1 Hygro between the HindIII and XbaI sites.

5

Transient Transfections: HEK293 cells were plated out in 6 well dishes (200,000 cells/well) and transiently transfected with NOT1 luciferase reporter and increasing amounts of either NOT1 or NOT1a expression plasmid (a construct expressing renilla luciferase was included as an internal transfection efficiency control). After 24-36 hours incubation cells were lysed and assayed for luciferase activity was measured using the Dual-Luciferase Reporter Assay System (Promega). Data are presented as fold induction after normalisation to renilla luciferase.

10

Results

Transfection of increasing amounts of full length NOT1 results in a dose dependent induction of luciferase activity (13-fold at 3µg). The activity of the NOT1a variant is markedly lower (2-fold induction observed at 3µg). The lower stimulatory activity of the NOT1a variant probably reflects the loss of the transcriptional activation domain AF2 .

15

All publications, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference as if each individual publication were specifically and individually indicated to be incorporated by reference herein as though fully set forth.

20

SEQUENCE INFORMATION**SEQ ID NO:1 (Full-length NOT1a cDNA)**

ATGCCTTGTGTTTCAGGCGCAGTATGGGTCTCGCCTCAAGGAGCCAGCCCCGCTTCTCAGAGCTACAGTTAC
 CACTCTTCGGGAGAATACAGCTCCGATTTCTTAACTCCAGAGTTTGTCAAGTTTAGCATGGACCTCACCAAC
 5 ACTGAAATCACTGCCACCACTTCTCTCCCCAGCTTCAGTACCTTTATGGACAACTACAGCACAGGCTACGAC
 GTCAAGCCACCTTGCTTGTACCAAATGCCCCGTGTCGGACAGCAGTCTCCATTAAGGTAGAAGACATTTCAG
 ATGCACAACCTACCAGCAACACAGCCACCTGCCCCCCCCAGTCTGAGGAGATGATGCCGCACTCCGGGTTCGGTT
 TACTACAAGCCCTCTCGCCCCGACGCCACCACCCGGGCTTCCAGGTGCAGCACAGCCCCATGTGGGAC
 GACCCGGGATCTCTCCACAACCTCCACCAGAACTACGTGGCCACTACGCACATGATCGAGCAGAGGAAAACG
 10 CCAGTCTCCGCGCTCTCCCTCTTCTCTTTAAGCAATCGCCCCCTGGCACCCCGGTGTCTAGTTGCCAGATG
 CGCTTCGACGGGCCCCCTGCACGTCCCCATGAACCCGGAGCCCGCCGGCAGCCACCACGTGGTGGACGGGCAG
 ACCTTCGCTGTGCCCAACCCCATTCGCAAGCCCGCGTCCATGGGCTTCCCGGGCCTGCAGATCGGCCACGCG
 TCTCAGCTGCTCGACACGCAGGTGCCCTCACCGCGTCGCGGGGCTCCCCCTCCAACGAGGGGCTGTGCGCT
 GTGTGTGGGGACAACGCGGCCTGCCAACACTACGGCGTGCACCTGTGAGGGCTGCAAAGGCTTCTTTAAG
 15 CGCACAGTGCAAAAAAATGCAAAATACGTGTGTTTAGCAAATAAAAACTGCCAGTGGACAAGCGTCGCCGG
 AATCGCTGTCACTGCTGCGGATTTTCAAGTGCCTGGCTGTTGGGATGGTCAAAGAAGTGGTTCGCACAGAC
 AGTTTAAAGGCCGGAGAGGTGCTTTGCCCTCGAAACCGAAGAGCCACAGGAGCCCTCTCCCCCTTCGCCC
 CCGGTGAGTCTGATCAGTGCCCTCGTCAGGGCCATGTGACTCCAACCCGGCTATGACCAGCCTGGACTAT
 TCCAGGTTCCAGGCGAACCTGACTATCAAATGAGTGGAGATGACACCCAGCATATCCAGCAATTCTATGAT
 20 CTCCTGACTGGCTCCATGGAGATCATCGGGGCTGGGCAGAGAAGATCCCTGGCTTCGCAGACCTGCCCAA
 GCCGACCAAGACCTGCTTTTGAATCAGCTTCTTAGAACTGTTTGTCTTCGATTAGCATACAGAATATGA
 ACATCGACATTTCTGCCTTCTCCTGCATTGCTGCCCTGGCTATGGTCAAGAGAGACACGGGCTCAAGGAAC
 CCAAGAGAGTGGAAGAACTGCAAAACAAGATTGTAAATTGTCTCAAAGACCACGTGACTTTCAACAATGGGG
 GGTGTAACCGCCCCAATTATTTGTCCAACTGTTGGGGAAGCTCCAGAACTTCGTACCCTTTGCACACAGG
 25 GGCTACAGCGCATTTTCTACCTGAAATTGGAAGACTTGGTGCCACCGCCAGCAATAATTGACAACTTTTTC
 TGGACACTTTACCTTTCTAA

SEQ ID NO:2 (NOT1a polypeptide)

MPCVQAQYQSSPQGASPASQSYSHSSGEYSSDFLTPEFVKFSMDLTNTEITATTSLSFSTFMDNYSTGYD
 30 VKPPCLYQMPLSGQSSIKVEDIQMHNYQHQSHLFPQSEEMPHSGSVYKPSPTPTTPGFQVQHSMPWD
 DPGSLHNFHQNYVATTHMIEQRKTPVSRSLFSFKQSPPGTPVSSCQMRFDGPLHVPMPNPEPAGSHHVVDGQ
 TFAVPNPPIRKPA SMGF PGLQIGHASQLLDTQVPSPPSRGSPSNEGLCAVCGDNAACQHYGVRTCEGCKGFFK
 RTVQKNKYVCLANKNCPVDKRRNRNRCQYCRFQKCLAVGMVKEVVRTDSLKGRRGRLPSKPKSPQEPSPPSP
 PVSLISALVRAHVDSNPAMTSLDYSRFQANPDYQMSGDDTQHIQQFYDLLTGSMEIIRGWAEKIPGFADLPK
 35 ADQDLLFESAFLELFVLRLAYRI

SEQ ID NO:3 (Full-length NOT1 cDNA)

ATGCCTTGTGTTTCAGGCGCAGTATGGGTCTCGCCTCAAGGAGCCAGCCCCGCTTCTCAGAGCTACAGTTAC
 CACTCTTCGGGAGAATACAGCTCCGATTTCTTAACTCCAGAGTTTGTCAAGTTTAGCATGGACCTCACCAAC
 40 ACTGAAATCACTGCCACCACTTCTCTCCCCAGCTTCAGTACCTTTATGGACAACTACAGCACAGGCTACGAC
 GTCAAGCCACCTTGCTTGTACCAAATGCCCTGTCCGGACAGCAGTCTCCATTAAGGTAGAAGACATTTCAG
 ATGCACAACCTACCAGCAACACAGCCACCTGCCCCCCCCAGTCTGAGGAGATGATGCCGCACTCCGGGTTCGGTT
 TACTACAAGCCCTCTCGCCCCGACGCCACCACCCGGGCTTCCAGGTGCAGCACAGCCCCATGTGGGAC
 GACCCGGGATCTCTCCACAACCTCCACCAGAACTACGTGGCCACTACGCACATGATCGAGCAGAGGAAAACG

CCAGTCTCCCGCCTCTCCCTCTTCTCCTTTAAGCAATCGCCCCCTGGCACCCCGGTGTCTAGTTGCCAGATG
CGCTTCGACGGGGCCCTGCACGTCCCCATGAACCCGGAGCCCGCCGGCAGCCACCACGTGGTGGACGGGCAG
ACCTTCGCTGTGCCCAACCCCATTCGCAAGCCCGCTCCATGGGCTTCCCGGGCCTGCAGATCGGCCACGCG
TCTCAGCTGCTCGACACGCAGGTGCCCTCACCGCCGTCGCGGGGCTCCCCCTCCAACGAGGGGCTGTGCGCT
5 GTGTGTGGGGACAACGCGGCCTGCCAACACTACGGCGTGCGCACCTGTGAGGGCTGCAAAGGCTTCTTTAAG
CGCACAGTGCAAAAAATGCAAAATACGTGTGTTTAGCAAATAAAAACTGCCAGTGGACAAGCGTCGCCGG
AATCGCTGTCACTAGTCCGATTTTCAAGAGTGCTGGCTGTTGGGATGGTCAAAGAAGTGGTTCGCACAGAC
AGTTTAAAAGGCCGGAGAGGTGTTTTGCCCTCGAAACCGAAGAGCCACAGGAGCCCTCTCCCCCTTCGCCC
CCGGTGAGTCTGATCAGTGCCCTCGTCAGGGCCCATGTGCACTCCAACCCGGCTATGACCAGCCTGGACTAT
10 TCCAGGTTCCAGGCGAACCTGACTATCAAATGAGTGGAGATGACACCCAGCATATCCAGCAATTCTATGAT
CTCTGACTGGCTCCATGGAGATCATCCGGGGCTGGGCAGAGAAGATCCCTGGCTTCGCAGACCTGCCCAA
GCCGACCAAGACCTGCTTTTGAATCAGCTTCTTAGAACTGTTTGTCTTCGATTAGCATACAGGTCCAAC
CCAGTGGAGGGTAAACTCATCTTTTGAATGGGGTGGTCTTGACAGGTTGCAATGCGTTCTGCGCTTTGGG
GAATGGATTGATTCCATTGTTGAATTCTCTCCAACCTTGCAAGATATGAACATCGACATTTCTGCTTCTCC
15 TGCATTGCTGCCCTGGCTATGGTTCACAGAGAGACACGGGCTCAAGGAACCAAGAGAGTGAAGAAGTCAA
AACAAGATTGTAAATTGTCTCAAAGACCACGTGACTTTCAACAATGGGGGGTTGAACCGCCCAATTATTTG
TCCAAACTGTTGGGGAAGCTCCAGAACTTCGTACCTTTGCACACAGGGGCTACAGCGCATTTCTACCTG
AAATTGGAAGACTTGGTGCCACCGCCAGCAATAATTGACAACTTTTCTGGACACTTTACCTTTCTAA

20 **SEQ ID NO:4 (NOT1 polypeptide)**

MPCVQAQYGSSPQGASPASQSYSHSSGEYSSDFLTPEFVKFSMDLTNTEITATTSLPSSFSTFMDNYSTGYD
VKPPCLYQMPLSGQQSSIKVEDIQMHNYQQHSHLPPQSEEMPHSGSVYKPPSSPPTPTTPGFQVQHS PMWD
DPGSLHNFHQNYVATTHMIEQRKTPVSRSLSFQSPPGTPVSSCQMRFDGPLHVPMNPEPAGSHHVVDGQ
TFAVPNPIRKPA SMGFPGLQIGHASQLLDQVPSPPSRGSPSNEGLCAVCGDNAACQHYGVRTCEGCKGFFK
25 RTVQKNAKYVCLANKNCPVDKRRNRNRCQYCRFQKCLAVGMVKEVVRTDSLKGRRGRLPSKPKSPQEPSPPSP
PVSLISALVRAHVDSNPAMTSLDYSRFQANPDYQMSGDDTQHIQQFYDLLTGSM EII RGWAEKIPGFADLPK
ADQDLLFESAFLELFVLRLAYRSNPVEGKLIFCNGVVLHRLQCVRGFGEWIDSIVEFSSNLQNMNIDISAFS
CIAALAMVTERHGLKEPKRVEELQNKIVNCLKDHVTFNNGGLNRPNYLSKLLGKLPELRTLCTQGLQRIFYL
KLEDLVPPP A I IDKLFLDTLPF

30

Claims

1. An isolated polypeptide comprising a polypeptide sequence which has at least 95% identity to the polypeptide sequence of SEQ ID NO:2.
- 5 2. The polypeptide as claimed in claim 1 comprising the polypeptide sequence of SEQ ID NO:2.
3. The isolated polypeptide of SEQ ID NO:2.
- 10 4. An isolated polynucleotide comprising a polynucleotide sequence encoding a polypeptide sequence which has at least 95% identity to a polypeptide of SEQ ID NO:2; or a polynucleotide sequence complementary to said isolated polynucleotide.
- 15 5. An isolated polynucleotide which comprises a polynucleotide sequence which has at least 95% identity to that of SEQ ID NO:1; or a polynucleotide sequence complementary to said isolated polynucleotide.
6. An isolated polynucleotide selected from:
 - (a) a polynucleotide comprising a polynucleotide sequence encoding the polypeptide of SEQ ID
 - 20 NO:2;
 - (b) a polynucleotide encoding the polypeptide of SEQ ID NO:2; and
 - (c) the polynucleotide of SEQ ID NO:1;or a nucleotide sequence complementary to said isolated polynucleotide over the entire length thereof.
- 25 7. An expression vector comprising a polynucleotide capable of producing a polypeptide of claim 1 when said expression system is present in a compatible host cell.
8. A host cell comprising the expression vector of claim 7 or a membrane thereof expressing the
- 30 polypeptide of claim 1.
9. A process for producing a polypeptide of claim 1 comprising the step of culturing a host cell of claim 8 under conditions sufficient for the production of said polypeptide and recovering the polypeptide from the culture medium.
- 35 10. An antibody immunospecific for the polypeptide of any one of claims 1 to 3.

11. A method for screening to identify compounds which stimulate or which inhibit the function of the polypeptide of claim 1 which comprises a method selected from the group consisting of:
- 5 (a) measuring the binding of a candidate compound to the polypeptide (or to the cells or membranes expressing the polypeptide) or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound;
 - (b) measuring the binding of a candidate compound to the polypeptide (or to the cells or membranes expressing the polypeptide) or a fusion protein thereof in the presense of a labeled competitor;
 - 10 (c) testing whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells or cell membranes expressing the polypeptide;
 - (d) mixing a candidate compound with a solution containing a polypeptide of claim 1, to form a mixture, measuring the activity of the polypeptide in the mixture, and comparing the activity of
15 the mixture to a standard; or
 - (e) detecting the effect of a candidate compound on the production of mRNA encoding said polypeptide and said polypeptide in cells, using for instance, an ELISA assay.
12. A process for diagnosing a disease or a susceptibility to a disease in a subject related to
20 expression or activity of the polypeptide of claim 1 in a subject comprising:
- (a) determining the presence or absence of a mutation in the nucleotide sequence encoding said polypeptide in the genome of said subject; and/or
 - (b) analyzing for the presence or amount of said polypeptide expression in a sample derived from said subject.

SEQUENCE LISTING

<110> SmithKline Beecham plc

<120> Novel Compounds

<130> GP30209

<160> 4

<170> FastSEQ for Windows Version 3.0

<210> 1

<211> 1676

<212> DNA

<213> Homo sapiens

<400> 1

atgccttggtg	ttcaggcgca	gtatgggtcc	tcgcctcaag	gagccagccc	cgcttctcag	60
agctacagtt	accactcttc	gggagaatac	agctccgatt	tcttaactcc	agagtttgtc	120
aagtttagca	tggacctcac	caacactgaa	atcactgcc	ccacttctct	ccccagcttc	180
agtaccttta	tggacaacta	cagcacaggc	tacgacgtca	agccaccttg	cttgtaccaa	240
atgccccgtg	ccggacagca	gtcctccatt	aaggtagaag	acattcagat	gcacaactac	300
cagcaacaca	gccacctgcc	ccccagttct	gaggagatga	tgccgcactc	cgggtcgggt	360
tactacaagc	cctcctcgcc	cccgacgccc	accaccccg	gcttccaggt	gcagcacagc	420
cccatgtggg	acgaccggg	atctctccac	aacttccacc	agaactacgt	ggccactacg	480
cacatgatcg	agcagaggaa	aacgccagtc	tccegcctct	ccctcttctc	ctttaagcaa	540
tcgccccctg	gcaccccggt	gtctagttgc	cagatgcgct	tcgacgggcc	cctgcacgtc	600
cccatgaacc	cggagccgc	cggcagccac	cacgtggtgg	acgggcagac	cttcgctgtg	660
cccaacccca	ttcgcaagcc	cgcgtccatg	ggcttcccg	gcctgcagat	cggccacgcg	720
tctcagctgc	tcgacacgca	ggtgccctca	ccgccgtcgc	ggggctcccc	ctccaacgag	780
gggctgtgcg	ctgtgtgtgg	ggacaacgcg	gcctgccaac	actacggcgt	gcgcacctgt	840
gagggtgca	aaggcttctt	taagcgcaca	gtgcaaaaa	atgcaaaata	cgrgtgttta	900
gcaaaataaaa	actgccaggt	ggacaagcgt	cggcgaatc	gctgtcagta	ctgccgattt	960
cagaagtgcc	tggctgttgg	gatggtcaaa	gaagtgttc	gcacagacag	tttaaaaggc	1020
cggagaggtc	gtttgccctc	gaaaccgaag	agccacagg	agccctctcc	cccttcgccc	1080
ccggtgagtc	tgatcagtc	cctcgtcagg	gccccatgct	actccaaccc	ggctatgacc	1140
agcctggact	attccaggtt	ccaggcgaac	cctgactatc	aaatgagtgg	agatgacacc	1200
cagcatatcc	agcaattcta	tgatctcctg	actggctcca	tggagatcat	ccggggctgg	1260
gcagagaaga	tccttggtt	cgcagacctg	cccaaagccg	accaagacct	gctttttgaa	1320
tcagctttct	tagaactgtt	tgtccttcga	ttagcataca	gaatatgaac	atcgacattt	1380

```

ctgccttctc ctgcattgct gccctggcta tggtcacaga gagacacggg ctcaaggaac 1440
ccaagagagt ggaagaactg caaaacaaga ttgtaaattg tctcaaagac cactgactt 1500
tcaacaatgg ggggttgaaac cgccccaatt atttgtccaa actgttgggg aagctcccag 1560
aacttcgtac cctttgcaca caggggctac agcgcathtt ctacctgaaa ttggaagact 1620
tggtgccacc gccagcaata attgacaaac ttttcctgga cactttacct ttctaa 1676

```

<210> 2

<211> 455

<212> PRT

<213> Homo sapiens

<400> 2

```

Met Pro Cys Val Gln Ala Gln Tyr Gly Ser Ser Pro Gln Gly Ala Ser
 1           5           10           15
Pro Ala Ser Gln Ser Tyr Ser Tyr His Ser Ser Gly Glu Tyr Ser Ser
          20           25           30
Asp Phe Leu Thr Pro Glu Phe Val Lys Phe Ser Met Asp Leu Thr Asn
        35           40           45
Thr Glu Ile Thr Ala Thr Thr Ser Leu Pro Ser Phe Ser Thr Phe Met
       50           55           60
Asp Asn Tyr Ser Thr Gly Tyr Asp Val Lys Pro Pro Cys Leu Tyr Gln
      65           70           75           80
Met Pro Leu Ser Gly Gln Gln Ser Ser Ile Lys Val Glu Asp Ile Gln
          85           90           95
Met His Asn Tyr Gln Gln His Ser His Leu Pro Pro Gln Ser Glu Glu
        100          105          110
Met Met Pro His Ser Gly Ser Val Tyr Tyr Lys Pro Ser Ser Pro Pro
       115          120          125
Thr Pro Thr Thr Pro Gly Phe Gln Val Gln His Ser Pro Met Trp Asp
      130          135          140
Asp Pro Gly Ser Leu His Asn Phe His Gln Asn Tyr Val Ala Thr Thr
     145          150          155          160
His Met Ile Glu Gln Arg Lys Thr Pro Val Ser Arg Leu Ser Leu Phe
          165          170          175
Ser Phe Lys Gln Ser Pro Pro Gly Thr Pro Val Ser Ser Cys Gln Met
          180          185          190
Arg Phe Asp Gly Pro Leu His Val Pro Met Asn Pro Glu Pro Ala Gly
         195          200          205
Ser His His Val Val Asp Gly Gln Thr Phe Ala Val Pro Asn Pro Ile
        210          215          220
Arg Lys Pro Ala Ser Met Gly Phe Pro Gly Leu Gln Ile Gly His Ala
       225          230          235          240
Ser Gln Leu Leu Asp Thr Gln Val Pro Ser Pro Pro Ser Arg Gly Ser

```

245 250 255
 Pro Ser Asn Glu Gly Leu Cys Ala Val Cys Gly Asp Asn Ala Ala Cys
 260 265 270
 Gln His Tyr Gly Val Arg Thr Cys Glu Gly Cys Lys Gly Phe Phe Lys
 275 280 285
 Arg Thr Val Gln Lys Asn Ala Lys Tyr Val Cys Leu Ala Asn Lys Asn
 290 295 300
 Cys Pro Val Asp Lys Arg Arg Arg Asn Arg Cys Gln Tyr Cys Arg Phe
 305 310 315 320
 Gln Lys Cys Leu Ala Val Gly Met Val Lys Glu Val Val Arg Thr Asp
 325 330 335
 Ser Leu Lys Gly Arg Arg Gly Arg Leu Pro Ser Lys Pro Lys Ser Pro
 340 345 350
 Gln Glu Pro Ser Pro Pro Ser Pro Pro Val Ser Leu Ile Ser Ala Leu
 355 360 365
 Val Arg Ala His Val Asp Ser Asn Pro Ala Met Thr Ser Leu Asp Tyr
 370 375 380
 Ser Arg Phe Gln Ala Asn Pro Asp Tyr Gln Met Ser Gly Asp Asp Thr
 385 390 395 400
 Gln His Ile Gln Gln Phe Tyr Asp Leu Leu Thr Gly Ser Met Glu Ile
 405 410 415
 Ile Arg Gly Trp Ala Glu Lys Ile Pro Gly Phe Ala Asp Leu Pro Lys
 420 425 430
 Ala Asp Gln Asp Leu Leu Phe Glu Ser Ala Phe Leu Glu Leu Phe Val
 435 440 445
 Leu Arg Leu Ala Tyr Arg Ile
 450 455

<210> 3

<211> 1797

<212> DNA

<213> Homo sapiens

<400> 3

atgccttgtg ttcaggcgca gtatgggtcc tcgcctcaag gagccagccc cgcttctcag	60
agctacagtt accactcttc gggagaatac agctccgatt tcttaactcc agagtttgtc	120
aagtttagca tggacctcac caacttgaa atcactgcca ccacttctct cccagcttc	180
agtaccttta tggacaacta cagcacaggc tacgacgtca agccaccttg cttgtaccaa	240
atgcccctgt cgggacagca gtcctccatt aaggtagaag acattcagat gcacaactac	300
cagcaacaca gccacctgcc cccccagtct gaggagatga tgccgcactc cgggtcgggt	360
tactacaagc cctcctcgcc cccgacgccc accacccggg gcttccaggt gcagcacagc	420
cccatgtggg acgacccggg atctctccac aacttcacc agaactacgt ggccactacg	480
cacatgatcg agcagaggaa aacgccagtc tcccgcctct ccctcttctc ctttaagcaa	540

```

tcgccccctg gcaccccggt gtctagttgc cagatgcgct tcgacggggcc cctgcacgtc      600
cccatgaacc cggagcccg cggcagccac cagtggtgg acgggcagac cttcgctgtg      660
cccaacccca ttgcaagcc cgcgtccatg ggcttcccgg gcctgcagat cggccacgcg      720
tctcagctgc tcgacacgca ggtgccccca ccgccgtcgc ggggctcccc ctccaacgag      780
gggctgtgcg ctgtgtgtgg ggacaacgcg gcctgccaac actacggcgt gcgcacctgt      840
gagggctgca aaggcttctt taagcgacac gtgcaaaaa atgcaaaata cgtgtgttta      900
gcaataaaaa actgcccagt ggacaagcgt cgccggaatc gctgtcagta ctgccgattt      960
cagaagtgcc tggctgttgg gatggtcaaa gaagtgggtc gcacagacag tttaaaaggc     1020
cggagaggtc gtttgccctc gaaaccgaag agcccacagg agccctctcc ccttcgccc     1080
ccggtgagtc tgatcagtc cctcgtcagg gcccatgtcg actccaaccc ggctatgacc     1140
agcctggact attccagggt ccaggcgaac cctgactatc aaatgagtgg agatgacacc     1200
cagcatatcc agcaattcta tgatctcctg actggtcca tggagatcat ccggggctgg     1260
gcagagaaga tccctggctt cgcagacctg cccaaagccg accaagacct gctttttgaa     1320
tcagctttct tagaactgtt tgtccttcga ttagcataca ggtccaaccc agtggagggt     1380
aaactcatct tttgcaatgg ggtggtcttg cacagggtgc aatgcgttcg tggctttggg     1440
gaatggattg attccattgt tgaattctcc tccaacttgc agaatatgaa catcgacatt     1500
tctgccttct cctgcattgc tgcctggct atggtcacag agagacacgg gctcaaggaa     1560
cccaagagag tggaagaact gcaaaacaag attgtaaatt gtctcaaaga ccacgtgact     1620
ttcaacaatg ggggggtgaa ccgccccaat tatttgtcca aactgttggg gaagctccca     1680
gaacttcgta cctttgcac acaggggcta cagcgcatth tctacctgaa attggaagac     1740
ttggtgccac cgccagcaat aattgacaaa ctttctctgg acactttacc tttctaa     1797

```

<210> 4

<211> 598

<212> PRT

<213> Homo sapiens

<400> 4

```

Met Pro Cys Val Gln Ala Gln Tyr Gly Ser Ser Pro Gln Gly Ala Ser
 1             5             10             15
Pro Ala Ser Gln Ser Tyr Ser Tyr His Ser Ser Gly Glu Tyr Ser Ser
      20             25             30
Asp Phe Leu Thr Pro Glu Phe Val Lys Phe Ser Met Asp Leu Thr Asn
      35             40             45
Thr Glu Ile Thr Ala Thr Thr Ser Leu Pro Ser Phe Ser Thr Phe Met
      50             55             60
Asp Asn Tyr Ser Thr Gly Tyr Asp Val Lys Pro Pro Cys Leu Tyr Gln
      65             70             75             80
Met Pro Leu Ser Gly Gln Gln Ser Ser Ile Lys Val Glu Asp Ile Gln
      85             90             95
Met His Asn Tyr Gln Gln His Ser His Leu Pro Pro Gln Ser Glu Glu
      100            105            110
Met Met Pro His Ser Gly Ser Val Tyr Tyr Lys Pro Ser Ser Pro Pro

```

```

115          120          125
Thr Pro Thr Thr Pro Gly Phe Gln Val Gln His Ser Pro Met Trp Asp
130          135          140
Asp Pro Gly Ser Leu His Asn Phe His Gln Asn Tyr Val Ala Thr Thr
145          150          155          160
His Met Ile Glu Gln Arg Lys Thr Pro Val Ser Arg Leu Ser Leu Phe
165          170          175
Ser Phe Lys Gln Ser Pro Pro Gly Thr Pro Val Ser Ser Cys Gln Met
180          185          190
Arg Phe Asp Gly Pro Leu His Val Pro Met Asn Pro Glu Pro Ala Gly
195          200          205
Ser His His Val Val Asp Gly Gln Thr Phe Ala Val Pro Asn Pro Ile
210          215          220
Arg Lys Pro Ala Ser Met Gly Phe Pro Gly Leu Gln Ile Gly His Ala
225          230          235          240
Ser Gln Leu Leu Asp Thr Gln Val Pro Ser Pro Pro Ser Arg Gly Ser
245          250          255
Pro Ser Asn Glu Gly Leu Cys Ala Val Cys Gly Asp Asn Ala Ala Cys
260          265          270
Gln His Tyr Gly Val Arg Thr Cys Glu Gly Cys Lys Gly Phe Phe Lys
275          280          285
Arg Thr Val Gln Lys Asn Ala Lys Tyr Val Cys Leu Ala Asn Lys Asn
290          295          300
Cys Pro Val Asp Lys Arg Arg Arg Asn Arg Cys Gln Tyr Cys Arg Phe
305          310          315          320
Gln Lys Cys Leu Ala Val Gly Met Val Lys Glu Val Val Arg Thr Asp
325          330          335
Ser Leu Lys Gly Arg Arg Gly Arg Leu Pro Ser Lys Pro Lys Ser Pro
340          345          350
Gln Glu Pro Ser Pro Pro Ser Pro Pro Val Ser Leu Ile Ser Ala Leu
355          360          365
Val Arg Ala His Val Asp Ser Asn Pro Ala Met Thr Ser Leu Asp Tyr
370          375          380
Ser Arg Phe Gln Ala Asn Pro Asp Tyr Gln Met Ser Gly Asp Asp Thr
385          390          395          400
Gln His Ile Gln Gln Phe Tyr Asp Leu Leu Thr Gly Ser Met Glu Ile
405          410          415
Ile Arg Gly Trp Ala Glu Lys Ile Pro Gly Phe Ala Asp Leu Pro Lys
420          425          430
Ala Asp Gln Asp Leu Leu Phe Glu Ser Ala Phe Leu Glu Leu Phe Val
435          440          445
Leu Arg Leu Ala Tyr Arg Ser Asn Pro Val Glu Gly Lys Leu Ile Phe
450          455          460

```


Cys Asn Gly Val Val Leu His Arg Leu Gln Cys Val Arg Gly Phe Gly
 465 470 475 480
 Glu Trp Ile Asp Ser Ile Val Glu Phe Ser Ser Asn Leu Gln Asn Met
 485 490 495
 Asn Ile Asp Ile Ser Ala Phe Ser Cys Ile Ala Ala Leu Ala Met Val
 500 505 510
 Thr Glu Arg His Gly Leu Lys Glu Pro Lys Arg Val Glu Glu Leu Gln
 515 520 525
 Asn Lys Ile Val Asn Cys Leu Lys Asp His Val Thr Phe Asn Asn Gly
 530 535 540
 Gly Leu Asn Arg Pro Asn Tyr Leu Ser Lys Leu Leu Gly Lys Leu Pro
 545 550 555 560
 Glu Leu Arg Thr Leu Cys Thr Gln Gly Leu Gln Arg Ile Phe Tyr Leu
 565 570 575
 Lys Leu Glu Asp Leu Val Pro Pro Pro Ala Ile Ile Asp Lys Leu Phe
 580 585 590
 Leu Asp Thr Leu Pro Phe
 595

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 00/02317

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C07K14/705 C12Q1/68 G01N33/50

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K C12Q G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, MEDLINE, STRAND, EMBL, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 94 04675 A (MAGES HANS WERNER ;KROCZEK RICHARD (DE)) 3 March 1994 (1994-03-03)	1,2,4-12
Y	the whole document page 11, line 12-20 page 11, line 28 -page 12, line 21 examples 3,4 claims 1,10	3
X	--- TORII T ET AL: "Organization of the human orphan nuclear receptor Nurrl gene" GENE, vol. 230, no. 2, 16 April 1999 (1999-04-16), pages 225-232, XP004165552	1,2,4-6
Y	abstract page 230, left-hand column, line 6 -page 231, left-hand column, line 7 figure 5	3
	--- -/-	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

28 September 2000

Date of mailing of the international search report

13/10/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

van de Kamp, M

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 00/02317

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>CASTILLO S O ET AL.: "A divergent role of COOH-terminal domains in Nurrl and Nur77 transactivation"</p> <p>GENE EXPRESSION,</p> <p>vol. 7, no. 1, 1998, pages 1-12,</p> <p>XP000946517</p> <p>abstract</p> <p>page 4, left-hand column, line 33</p> <p>-right-hand column, line 46</p> <p>figure 1</p> <p>---</p>	1,4
X	<p>DATABASE EMROD</p> <p>EMBL;</p> <p>ID MMU86783, AC U86783,</p> <p>29 September 1997 (1997-09-29)</p> <p>CASTILLO S O ET AL.: "Mus musculus steroid/thyroid hormone orphan nuclear receptor (Nurrl) gene, complete cds"</p> <p>XP002148027</p> <p>page 1, line 54,55</p> <p>page 2, line 21-35</p> <p>page 3, line 24</p>	1,4
X	<p>-& CASTILLO S O ET AL.: "Organization, sequence, chromosomal localization, and promoter identification of the mouse orphan nuclear receptor Nurrl gene."</p> <p>GENOMICS,</p> <p>vol. 41, no. 2,</p> <p>15 April 1997 (1997-04-15), pages 250-257,</p> <p>XP000946433</p> <p>abstract</p> <p>figure 1</p> <p>page 252, right-hand column, line 14-22</p> <p>---</p>	1,4
X	<p>DATABASE SWISSPROT 'Online!</p> <p>STRAND;</p> <p>ID NR42_MOUSE, AC 006219,</p> <p>1 June 1994 (1994-06-01)</p> <p>LAW S W ET AL.: "Orphan nuclear receptor Nurrl (Nur-related factor 1)"</p> <p>XP002148028</p> <p>page 1, line 29,30</p> <p>page 2, line 13,14</p>	1,4
X	<p>-& LAW S W ET AL.: "Identification of a new brain-specific transcription factor, NURR1"</p> <p>MOLECULAR ENDOCRINOLOGY,</p> <p>vol. 6, no. 12, December 1992 (1992-12),</p> <p>pages 2129-2135, XP000946429</p> <p>abstract</p> <p>---</p> <p>-/--</p>	1,4

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 00/02317

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 26063 A (MAIRA MARIO ;DROUIN JACQUES (CA); MONTREAL INST RECH CLINIQUES (CA) 18 June 1998 (1998-06-18) page 3, line 3-24 page 6, line 8-12 example 22 claims 14-16	11
A	--- MARUYAMA K ET AL.: "The NGFI-B subfamily of the nuclear receptor superfamily (review)" INTERNATIONAL JOURNAL OF ONCOLOGY, vol. 12, no. 6, June 1998 (1998-06), pages 1237-1243, XP000946431 abstract	1,4
A	--- STADEL J M ET AL: "Orphan G protein-coupled receptors: a neglected opportunity for pioneer drug discovery" TRENDS IN PHARMACOLOGICAL SCIENCES, vol. 18, no. 11, 1 November 1997 (1997-11-01), pages 430-437, XP004099345 abstract	11
P,X	--- SALIN-NORDSTROM, T. ET AL.: "Function of human not (Nurr1 /NR4A2) splice variants in development and pathologies." SOCIETY FOR NEUROSCIENCE ABSTRACTS. MEETING INFO.: 29TH ANNUAL MEETING OF THE SOCIETY FOR NEUROSCIENCE. MIAMI BEACH, FLORIDA, USA OCTOBER 23-28, 1999 SOCIETY FOR NEUROSCIENCE., vol. 25, no. 1-2, 23 October 1999 (1999-10-23), page 1754 XP000949970 abstract	1-12
T	--- KILPATRICK G J ET AL: "7TM receptors: the splicing on the cake" TRENDS IN PHARMACOLOGICAL SCIENCES, vol. 20, no. 7, 1 July 1999 (1999-07-01), pages 294-301, XP004171850 abstract -----	1-6

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/GB 00/02317

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9404675 A	03-03-1994	AU 4950393 A	15-03-1994
WO 9826063 A	18-06-1998	CA 2192754 A	12-06-1998
		AU 5470598 A	03-07-1998

Form PCT/ISA/210 (patent family annex) (July 1992)

SEQUENCE LISTING

<110> SmithKline Beecham plc

<120> Novel Compounds

<130> GP30209

<160> 4

<170> PasteBQ for Windows Version 3.0

<210> 1

<211> 1676

<212> DNA

<213> Homo sapiens

<400> 1

atgcttctgtg tttagggcga gtatgggtcc tgcctcag gagccagccc cgtttctcag	60
agctacagtt accactcttc gggagaatac agctcggatt tcttaactcc agagtctgtc	120
aagtcttagca tggacctcac caacactgaa atcaactgca ccactctctt cccagcttc	180
agtaccttta tggacaacta cggacaggc taagacgtca agccaccttg ctgtaccac	240
atgcccctgt ccggacagca gtccctccatt aaggtagaag acattcagat gcacaactac	300
cagcaacaca ggcacctgac ccccaagtct gaggagatga tgcggacctc cgggkcggtt	360
tactacaagc cctcctogcc cccgaagccc accacccctg gcttcagggt gcagcacagc	420
cccatgtggg acgaaccctg alctctccac aacttccacn agaactacgt ggccactacg	480
cacatgatog agcagaggga aacggcagtc tcccgcctct cctctctctt ctttaagosa	540
tgcacctctg gacccccgtt gtctagttag cagatgcgtt tcyacgggac cctgcacgtc	600
cccatgaacc cggagccctg cggcagccac caogtggtag aagggcagac cttcgctgtg	660
ccccaccca ttgcgaagcc cgggteratg ggttccctgg gctgacagat cggccacgog	720
tctcagctgc tcyacacgca ggtgacctc ccgacctgyc ggggctcccc ctccaacgag	780
gggctgtgog ctgtgtgttg ggaacaaogcg gcttcccaac actacygctt gcgcacctgt	840
gagggctgca aaggcttctt ttagggcaca gtycaaaaa atgcaaaata cgtgtgttta	900
gcaaataaaa actgccagtt ggaacaaggt cggcgggaac gctgtcagta ctgongattt	960
cagaagtgc tggctgttg gattgtcaaa gaagtgggtc gnaacagacg tttaaaaggc	1020
cggagagggtc gtttgccttc gaacccgaag agcccacagg agccctctcc ccttccgctc	1080
cgggtgagtc tgatcagtg cctcgtcagg gccctgtcty actccacccc ggtatgacc	1140
agcctggact attcagggtt ccaggcgaac cctgactatc aaatgagttg agatgacac	1200
cagcatatcc agcaattctt tgatctcttg actggttcca tggagatcat cgggggcttg	1260
gcagaganga tccctggctt cgcagacctg cccaaagctg accaagacct gttttttgac	1320
tacgtttctt tagaactgtt tgtctctga ttagcatana gaatatgaac atcgacctt	1380

```

ctgccttctc ctgcattgct gccctaggcta tggtcacaga gagacacggy ctcaagggaac 1440
ccaagagagt ggaagaactg caaaacaaga ttgtaaattg tctcaagac cagctgactt 1500
tcaacaatgg ggggttgaa cgcctcaatt attgtccaa actgttggg aagctccag 1560
aacttcgtac cctttgcaca caggggctac aggcatttt ctactgaaa ttggaagact 1620
tggtgcaacc gccagcaata attgacaaac ttttctgga cctttacct ttctaa 1676

```

<210> 2

<211> 455

<212> PRT

<213> Homo sapiens

<400> 2

```

Met Pro Cys Val Gln Ala Gln Tyr Gly Ser Ser Pro Gln Gly Ala Ser
1          5          10          15
Pro Ala Ser Gln Ser Tyr Ser Tyr His Ser Ser Gly Glu Tyr Ser Ser
20          25          30
Asp Phe Leu Thr Pro Glu Phe Val Lys Phe Ser Met Asp Leu Thr Asn
35          40          45
Thr Glu Ile Thr Ala Thr Thr Ser Leu Pro Ser Phe Ser Thr Phe Met
50          55          60
Asp Asn Tyr Ser Thr Gly Tyr Asp Val Lys Pro Pro Cys Leu Tyr Gln
65          70          75          80
Met Pro Leu Ser Gly Gln Gln Ser Ser Ile Lys Val Glu Asp Ile Gln
85          90          95
Met His Asn Tyr Gln Gln His Ser His Leu Pro Pro Gln Ser Glu Glu
100         105         110
Met Met Pro His Ser Gly Ser Val Tyr Tyr Lys Pro Ser Ser Pro Pro
115         120         125
Thr Pro Thr Thr Pro Gly Phe Gln Val Gln His Ser Pro Met Trp Asp
130         135         140
Asp Pro Gly Ser Leu His Asn Phe His Gln Asn Tyr Val Ala Thr Thr
145         150         155         160
His Met Ile Glu Gln Arg Lys Thr Pro Val Ser Arg Leu Ser Leu Phe
165         170         175
Ser Phe Lys Gln Ser Pro Pro Gly Thr Pro Val Ser Ser Cys Gln Met
180         185         190
Arg Phe Asp Gly Pro Leu His Val Pro Met Asn Pro Glu Pro Ala Gly
195         200         205
Ser His His Val Val Asp Gly Gln Thr Phe Ala Val Pro Asn Pro Ile
210         215         220
Arg Lys Pro Ala Ser Met Gly Phe Pro Gly Leu Gln Ile Gly His Ala
225         230         235         240
Ser Gln Leu Leu Asp Thr Gln Val Pro Ser Pro Pro Ser Arg Gly Ser

```

245 250 255
 Pro Ser Asn Glu Gly Leu Cys Ala Val Cys Gly Asp Asn Ala Ala Cys
 260 265 270
 Gln His Tyr Gly Val Arg Thr Cys Glu Gly Cys Lys Gly Phe Phe Lys
 275 280 285
 Arg Thr Val Gln Lys Asn Ala Lys Tyr Val Cys Leu Ala Asn Lys Asn
 290 295 300
 Cys Pro Val Asp Lys Arg Arg Arg Asn Arg Cys Gln Tyr Cys Arg Phe
 305 310 315 320
 Gln Lys Cys Leu Ala Val Gly Met Val Lys Glu Val Val Arg Thr Asp
 325 330 335
 Ser Leu Lys Gly Arg Arg Gly Arg Leu Pro Ser Lys Pro Lys Ser Pro
 340 345 350
 Gln Glu Pro Ser Pro Pro Ser Pro Pro Val Ser Leu Ile Ser Ala Leu
 355 360 365
 Val Arg Ala His Val Asp Ser Asn Pro Ala Met Thr Ser Leu Asp Tyr
 370 375 380
 Ser Arg Phe Gln Ala Asn Pro Asp Tyr Gln Met Ser Gly Asp Asp Thr
 385 390 395 400
 Gln His Ile Gln Gln Phe Tyr Asp Leu Leu Thr Gly Ser Met Glu Ile
 405 410 415
 Ile Arg Gly Trp Ala Glu Lys Ile Pro Gly Phe Ala Asp Leu Pro Lys
 420 425 430
 Ala Asp Gln Asp Leu Leu Phe Glu Ser Ala Phe Leu Glu Leu Phe Val
 435 440 445
 Leu Arg Leu Ala Tyr Arg Ile
 450 455

<210> 3

<211> 1797

<212> DNA

<213> Homo sapiens

<400> 3

atgccttggtg ttccaggogca gstatgggtcc tgcctcaag gagccagccc cgtttctcag 60
 agctacagtt accactcttc gggagatcac agctccgalt tcttaactcc agagtttgtc 120
 aagtttagta tggacttacc caacactgaa atcactgcca ccacttctct cccagcttc 180
 agtaccttta tggacaacta cagcacaggc tacyaogtca agccacottg ctctglaccan 240
 atgcctctgt ccggacagca gtctccatt aaggtagaa acattacat gccacactac 300
 cagcaacaca gccacatgca ccccaagtct gaggagatga tgcgcacac cgggtcggtt 360
 taatacaagc cctctctgccc ccgacgccc accaccccg gcttcaggt gcagcacagc 420
 cccatgtggg agcaccggg atctctccac aacttccac agaactacgt ggcactacg 480
 cacatgateg agcagagga aacgccagtc tcccgctct cctcttctc cttaagcga 540


```

tcgccccctg gcaccccggg gtctagttgc cagatgcgct togaaggggc cctgcaogtc      600
cccatgaacc cggagcccg cggcagccac cagctgggtg aogggcagac cctogctgtg      650
ccccacccca ttcgcaagcc cgcgtccatg ggcttccegg gectgcagut cggccacgeg      720
tctcagctgc togaacagca ggtgccctca cggcgtcgc ggggctcccc ntcnaangag      780
gggctgtgog ctgtgtgtgg ggataacgog gcctgccaac actaoggggt gcgcacctgt      840
gagggctgca aaggcttctt taagcgacac gtgcacaaaa atgcacaaata ogtgtgttta      900
gcaataaana actgccctag ggacaagcgt cgcgggaatc gctgtcagtc ctgccgattt      960
cagaagtgcg tggctgttgg gatggtaaaa gaagtgggtc gcacagacag tttaaaaggc     1020
cggagagggt gtttgccctc gaaacccaag agcccacagg agccctctcc ccttcgccc     1080
cgggtgagtc tcatcagtc cctcgtcagg gcccatgtcg actccaaccc ggctatgacc     1140
agcctggact attccagggt ccaggcgac cctgactatc aaatgagtg agelyacacc     1200
cagcatatcc agcaattcta tcatctctct actggtctca tggggtatct cgggggctgg     1260
gcagagaaga tccctggctt cgcagacctg cccaaagccg accaagacct gctttttgaa     1320
tcagctttct cagaactgtt tgtctctga ttagctatac ggtccacccc agtggagggt     1380
aaactctct tttgcattgg ggtggtcttg cacagggtgc natggtctcg tggctttggg     1440
gaatggattg attccattgt tgaattcttc tccaacttgc agaatatgaa catcgacatt     1500
cttgccttct cctgatttgc tgccttggct atggtcacag agagacacgg gctcaaggaa     1560
cccagagag tggaggaact gcaaacacag attgtaatc gtctcaaga ccacgtgact     1620
ttcaacaatg ggggggtgaa ccgcaccaat tattgtcca aactgttggg gaagctccca     1680
gaacttcgta ccttttgcac acaggggcta cagcgattt tctacotgaa attggaagac     1740
ttggtgccac cggcagcaat aattgacaaa ctttctctgg acactttacc ttctaa      1797

```

<210> 4

<211> 598

<212> PRT

<213> Homo sapiens

<400> 4

```

Met Pro Cys Val Gln Ala Gln Tyr Gly Ser Ser Pro Gln Gln Ala Ser
 1              5              10              15
Pro Ala Ser Gln Ser Tyr Ser Tyr His Ser Ser Gly Glu Tyr Ser Ser
      20              25              30
Asp Phe Leu Thr Pro Gln Phe Val Lys Phe Ser Met Asp Leu Thr Asn
      35              40              45
Thr Glu Ile Thr Ala Thr Thr Ser Leu Pro Ser Phe Ser Thr Phe Met
      50              55              60
Asp Asn Tyr Ser Thr Gly Tyr Asp Val Lys Pro Pro Cys Leu Tyr Gln
      65              70              75              80
Met Pro Leu Ser Gly Gln Gln Ser Ser Ile Lys Val Glu Asp Ile Gln
      85              90              95
Met His Asn Tyr Gln Gln His Ser His Leu Pro Pro Gln Ser Glu Glu
      100             105             110
Met Met Pro His Ser Gly S r Val Tyr Tyr Lys Pro Ser Ser Pro Pro

```

115	120	125
Thr Pro Thr Thr Pro Gly Phe Gln Val Gln His Ser Pro Met Trp Asp		
130	135	140
Asp Pro Gly Ser Leu His Asn Phe His Gln Asn Tyr Val Ala Thr Thr		
145	150	155
His Met Ile Gln Gln Arg Lys Thr Pro Val Ser Arg Leu Ser Leu Phe		160
	165	170
Ser Phe Lys Gln Ser Pro Pro Gly Thr Pro Val Ser Ser Cys Gln Met		175
	180	185
Arg Phe Asp Gly Pro Leu His Val Pro Met Asn Pro Gln Pro Ala Gly		190
	195	200
Ser His His Val Val Asp Gly Gln Thr Phe Ala Val Pro Asn Pro Ile		205
	210	215
Arg Lys Pro Ala Ser Met Gly Phe Pro Gly Leu Gln Ile Gly His Ala		220
225	230	235
Ser Gln Leu Leu Asp Thr Gln Val Pro Ser Pro Pro Ser Arg Gly Ser		240
	245	250
Pro Ser Asn Gln Gly Leu Cys Ala Val Cys Gly Asp Asn Ala Ala Cys		255
	260	265
Gln His Tyr Gly Val Arg Thr Cys Gln Gly Cys Lys Gly Phe Phe Lys		270
	275	280
Arg Thr Val Gln Lys Asn Ala Lys Tyr Val Cys Leu Ala Asn Lys Asn		285
	290	295
Cys Pro Val Asp Lys Arg Arg Arg Asn Arg Cys Gln Tyr Cys Arg Phe		300
305	310	315
Gln Lys Cys Leu Ala Val Gly Met Val Lys Gln Val Val Arg Thr Asp		320
	325	330
Ser Leu Lys Gly Arg Arg Gly Arg Leu Pro Ser Lys Pro Lys Ser Pro		335
	340	345
Gln Gln Pro Ser Pro Pro Ser Pro Pro Val Ser Leu Ile Ser Ala Leu		350
	355	360
Val Arg Ala His Val Asp Ser Asn Pro Ala Met Thr Ser Leu Asp Tyr		365
	370	375
Ser Arg Phe Gln Ala Asn Pro Asp Tyr Gln Met Ser Gly Asp Asp Thr		380
385	390	395
Gln His Ile Gln Gln Phe Tyr Asp Leu Leu Thr Gly Ser Met Gln Ile		400
	405	410
Ile Arg Gly Trp Ala Gln Lys Ile Pro Gly Phe Ala Asp Leu Pro Lys		415
	420	425
Ala Asp Gln Asp Leu Leu Phe Gln Ser Ala Phe Leu Gln Leu Ph Val		430
	435	440
Leu Arg Leu Ala Tyr Arg Ser Asn Pro Val Gln Gly Lys Leu Ile Phe		445
	450	455
		460

Leu His Arg Leu Gln Cys Val Arg Gly Phe Gly
 470 475 480
 Ile Val Glu Phe Ser Ser Asn Leu Gln Asn Met
 490 495
 Ala Phe Ser Cys Ile Ala Ala Leu Ala Met Val
 505 510
 Leu Lys Glu Pro Lys Arg Val Glu Glu Leu Gln
 515 520 525
 Asn Lys Ile Val Asn Cys Leu Lys Asp His Val Thr Phe Asn Asn Gly
 530 535 540
 Gly Leu Asn Arg Pro Asn Tyr Leu Ser Lys Leu Leu Gly Lys Leu Pro
 545 550 555 560
 Glu Leu Arg Thr Leu Cys Thr Gln Gly Leu Gln Arg Ile Phe Tyr Leu
 565 570 575
 Lys Leu Glu Asp Leu Val Pro Pro Pro Ala Ile Ile Asp Lys Leu Phe
 580 585 590
 Leu Asp Thr Leu Pro Phe
 595